

学校编码: 10384

分类号_____密级_____

学号: X2013232178

UDC_____

厦门大学

工程硕士学位论文

远程教育模式下智能答疑系统的
设计与实现

Design and Implementation of Intelligent Question
Answering System in Distance Education Mode

林珊

指导教师: 王美红助理教授

专业名称: 软件工程

论文提交日期: 2016 年 9 月

论文答辩日期: 2016 年 11 月

学位授予日期: 2016 年 12 月

指导教师: _____

答辩委员会主席: _____

2016 年 9 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下，独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（
）课题（组）
的研究成果，获得（
）课题（组）经费或实验室的
资助，在（
）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ） 1.经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

（ ☒ ） 2.不保密，适用上述授权。

（请在以上相应括号内打。√。或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

摘要

互联网技术的发展推动远程教育模式应运而生,通过有机结合互联网技术和计算机辅助教学技术进行教学改革,使得远程教育模式渐成现代教育发展新趋势。智能答疑作为远程教育系统的重要部分,是用来评估远程教育系统优劣性的重要指标,其打破传统的时空限制,进行在线实时答疑。智能答疑综合运用自然语言处理、信息检索等多学科知识,为教师与学生间的交流互动提供了保证,对素质教育的实施及教学改革的推进有一定裨益。

本文首先介绍系统开发的背景、目的和意义,并提出智能答疑系统的解决方案。其次,详细分析系统的用户需求、功能需求和非功能需求,构建完善的体系结构,设计数据库表结构。接着,详细研究在智能答疑模块中如何有效应用中文信息处理技术,比如自动分词技术、句子语义相似度计算等。最后,为提高 FAQ 库的查询效率,本文进一步研究数据库快速定位技术,通过比较正排索引结构和倒排索引结构,结合智能答疑系统所涉及的数据特点,最终决定采用倒排索引结构来进行 FAQ 库的检索,提高关键词检索的效率和精确度。

本文设计并开发的智能答疑系统是搭建在微软公司开发的 ASP.NET 平台上,在开始开发系统功能模块前,先配置 ASP.NET 开发环境,为 WEB 项目的开发做足准备。研究 MVC 设计模式和 ASP.NET MVC 框架,采用浏览器/服务器体系结构进行智能答疑系统各模块的设计和实现,并对智能答疑系统进行详细的测试,编写测试用例,保证系统达到需求和系统设计的要求。

关键词: 中文分词; 智能答疑; 远程教育

Abstract

With the development of Internet technology, distance-education mode, which effectively combines Internet technology and computer aided instruction to reform teaching mode, comes into being. It has gradually promoted the distance-education mode to be a new trend of the development of modern education. As a key part in distance-education mode, intelligent question answering, which is an important factor used to evaluate the robustness of a distance education system, breaks time and space limitations which exist in traditional question answering way and it deals with online and real-time answering. Intelligent-answering involves various aspects of multi-disciplinary knowledge and comprehensively uses natural language processing and information retrieval, providing high-quality interaction between teachers and students and playing a positive role in promoting the implementation of quality education and teaching reform.

First of all, we clarify the research background, purpose and significance and propose the solution of intelligent-answering system. Second, we analyze user requirements, functional requirements and non-functional requirements in detail, building complete architecture and designing the database structure further. Third, we study how to apply the Chinese information processing in intelligent-answering system, like Chinese word segmentation and sentence semantic similarity computation.

In addition, to effectively improve efficiency of query FAQ library, this dissertation further researches quickly positioning technology. Contrast index and inverted index and a detailed analysis of data characteristics, this dissertation finally decides to take the inverted index structure to improve accuracy and efficiency of keyword search.

In this dissertation, the IQAS is built on the platform .NET, based on MVC design pattern and adopting B/S architecture for the development, implement and detailed tests of the functional modules in the system.

Key Words: Chinese Word Segmentation; Intelligent Question Answering; Distance Education

厦门大学博士论文摘要库

目录

第 1 章 引言	1
1.1 研究背景	1
1.2 研究目的和意义	3
1.2.1 研究目的	3
1.2.2 研究意义	3
1.3 远程教育系统研究概述	3
1.3.1 发展历程和现状	3
1.3.2 系统的特点	4
1.4 智能答疑系统研究概述	6
1.4.1 概述	6
1.4.2 研究现状	6
1.5 本文主要研究内容	8
1.6 论文组织结构	9
第 2 章 开发平台及相关技术介绍	10
2.1 开发平台	10
2.1.1 Visual Studio 2010	10
2.1.2 Sql Server 2008	10
2.1.3 IIS 7.0	11
2.2 相关技术	11
2.2.1 B/S 体系结构	11
2.2.2 C#	12
2.2.3 ASP.NET	12
2.2.4 MVC 设计模式	13
2.2.5 ASP.NET MVC	15
2.2.6 LINQ To SQL 技术	16
2.3 本章小结	17
第 3 章 需求分析	18

3.1 系统目标	18
3.2 用户需求分析	18
3.3 系统功能需求分析	20
3.4 系统非功能需求分析	22
3.5 系统关键技术需求分析	22
3.6 本章小结	23
第 4 章 系统设计	24
4.1 系统体系结构	24
4.2 登录模块设计	26
4.3 智能答疑模块设计	27
4.3.1 快速中文分词机制.....	27
4.3.2 基于向量空间模型的句子相似度计算.....	30
4.3.3 基于《同义词词林》的语义相似度计算.....	32
4.3.4 索引结构.....	34
4.4 数据库设计	37
4.4.1 逻辑关系图（E-R 图）	37
4.4.2 数据库表结构设计.....	37
4.5 本章小结	39
第 5 章 系统实现与测试.....	40
5.1 系统三层体系结构的实现	40
5.1.1 Model 部分的实现	40
5.1.2 View 部分的实现	41
5.1.3 Controller 部分的实现	42
5.2 登录模块实现	44
5.3 智能答疑模块实现	46
5.4 系统测试	53
5.4.1 中文分词测试.....	53
5.4.2 中文句子相似度测试.....	54
5.5 本章小结	55

第 6 章 总结与展望	56
6.1 总结	56
6.2 展望	56
参考文献	57
致谢	59

Contents

Chapter 1 Introduction	1
1.1 Research Background	1
1.2 Research Purpose and Significance	3
1.2.1 Research Purpose	3
1.2.2 Research Significance	3
1.3 Research Overview of Tele-Education System	3
1.3.1 Development and Situation	3
1.3.2 Characteristics of System	4
1.4 Research Overview of Intelligent Question Answering System	6
1.4.1 Research Overview	6
1.4.2 Research Situation	6
1.5 Main Research Contents	8
1.6 Dissertation Structure	9
Chapter 2 Development Platform and Related Technologies	10
2.1 Development Platform	10
2.1.1 Visual Studio 2010	10
2.1.2 Sql Server 2008	10
2.1.3 IIS 7.0	11
2.2 Related Technologies	11
2.2.1 B/S Architecture	11
2.2.2 C#	12
2.2.3 ASP.NET	12
2.2.4 MVC Design Pattern	13
2.2.5 ASP.NET MVC	15
2.2.6 LINQ To SQL	16
2.3 Summary	17
Chapter 3 Requirements Analysis	18

3.1 System Objective	18
3.2 User Requirements Analysis	18
3.3 Functional Requirements Analysis	20
3.4 Non-Functional Requirements Analysis	22
3.5 Key Technologies Requirements Analysis.....	22
3.6 Summary.....	23
Chapter 4 Systematic Design	24
4.1 Architecture Structure	24
4.2 Login Module Design.....	26
4.3 Intelligent Question Answering Module Design.....	27
4.3.1 Fast Chinese Word Segmentation	27
4.3.2 Sentence Similarity Computation Based on Vector Space Model	30
4.3.3 Semantic Similarity Computation Based on Chinese Thesaurus-Tongyici Cilin.....	32
4.3.4 Index Structure.....	34
4.4 Database Design	37
4.4.1 Entity-Relation Diagrams	37
4.4.2 Database Tables' Structure Design	37
4.5 Summary.....	39
Chapter 5 System Implementation and Test	40
5.1 Implementation of Three-Layer Architecture.....	40
5.1.1 Implementation of Model	40
5.1.2 Implementation of View	41
5.1.3 Implementation of Controller	42
5.2 Implementation of Login Module.....	44
5.3 Implementation of Intelligent Question Answering Module.....	46
5.4 System Test	53
5.4.1 Chinese Word Segmentation Test	53
5.4.2 Chinese Sentence Similarity Test.....	54

5.5 Summary.....	55
Chapter 6 Conclusion and Prospect.	56
6.1 Conclusion	56
6.2 Prospect.....	56
References	57
Acknowledgement.	59

第1章 引言

1.1 研究背景

互联网诞生大大推动了软硬件技术、网络技术等计算机技术的快速发展,尤其是“互联网+”概念的提出,使得互联网在教育领域方面的应用愈加受到关注,远程教育应运而生,逐渐成为提高国民素质、普及文化素养的主阵地^[1]。远程教育模式打破传统束缚,摆脱教育领域在时间、空间上的局限,为教师和学生提供了双重便利。远程教育系统中一个重要的模块就是在线智能答疑模块,其设计的优劣是评估一个远程教育系统好坏以及等次最重要的指标^[2]。在线智能答疑模块综合应用了多个学科的专业知识,包括中文分词技术、信息检索技术、句子相似度分析技术、语义相似度分析技术等,通过这些技术对问题进行在线分析,并自动检索问题库给出匹配度最高的答案。其效率极高、操作极简,大大提高了远程教育系统的质量,提高了师生间的互动效率,保证远程教育能够高效有序地运行,为教育发展和教学改革增添光彩。

虽然远程教育模式是一种全新的并具有广阔前景的教学模式,但其诞生发展至今并不是一帆风顺的,其依然存在很多关键性的难题亟待解决。首先是时间差异性,师生都有很多的私人时间,这些私人时间在绝大多数情况下都是不交叉的,学生不可能占用教师私人时间来提问,教师也不可能占用学生私人时间来解答,这样导致答疑的时效滞后性,学生只好花费更多的时间另辟蹊径去解决难题;其次是空间差异性,由于地理空间上的局限,有时师生间可能相隔数千里,教师无法第一时间为学生提供面对面的答疑解惑。然而,研究表明,当学生面对难题,能在第一时间获得解答是非常有必要的,这样才能增强学生对所遇难题的印象,所以在远程教育系统中在线智能答疑模块已然成为不可缺少的关键性模块,尤其是在这种时空冲突的情况下,更是必不可少。通过阅读相关中外大量文献发现,不管是国内还是国外,大部分现有的远程教育系统设计在线智能答疑模块时都有一些考虑不是很到位、不是很周到的地方,总结归纳国内外现有的答疑模块,可以分为以下几种类型^[3-5]。

(1) 自主灵活、功能完善。这类答疑模块自带齐全的词库、句库等跟智能答疑模块相关的信息数据库,学生在遇到疑问后向系统提交问题,智能答疑模块自动解析问题,提取出问题中的关键词、关键句,并与已有的、存储在系统端

的数据库进行相似度匹配,将所匹配到问题的答案返回给学生用户,并将问题及答案更新至相应的词库、句库中。当然,如果在系统数据库中匹配不到相似的问题和答案,系统会自动将学生提交的问题上传给专家,等待专家对学生问题作出相应的回答。这类智能答疑模块综合应用了计算机领域中多个相关子领域的技术,包括数据挖掘、中文分词、动态检索等。

(2) 弱智能化,功能单一。这类答疑模块没在本地构建强大的词库、句库等相关的信息数据库,学生在遇到疑问后,无法第一时间获得有帮助的解答,显现出系统高度的弱智能化,功能单一。这类答疑模块需要基于强大的人力支撑,系统管理员或教师需要 24 小时无间断处于在线状态,才能在第一时间为学生答疑解惑,对时间有强烈的束缚。这类答疑模块模式有点像网络贴吧论坛的讨论模式,局限性强,且目前市场上绝大多数的答疑模块都属于这类。

(3) 弱电子化,功能缺失。这类答疑模块模式比较呆板,师生之间交流互动的方式仅限于系统公告、留言板、短信等,具有严重的滞后性。学生在遇到疑问后,无法第一时间获得有帮助的解答,使得许多教学信息、教学情况以及教学效果不能得到及时反馈。当然,虽说这类答疑模块已逐渐退出市场,但不可否认,它还是存在于一些网站、论坛、贴吧中的。

从以上文献资料中归纳的现有答疑模块可以发现,不管是国内还是国外,大部分现有的在线答疑模块都属于以上其中一种,详细分析这些答疑模块发现,这些答疑模块存在一些共性的缺点亟待解决,主要表现为:

(1) 答疑模块智能化程度低。以上介绍的几种答疑模块智能化程度不一,有的可以一定程度地完成自动答疑,有的可以完成人工前提下的在线答疑,而有时在智能性和实时性方面都有缺失,总而言之,都不同程度地缺乏一定智能性。而且,智能答疑模块解析语句、关键词的方式相对比较简单,系统对用户的计算机操作基础要求高。同时,用户对系统的评价渠道不畅通,不能及时地对系统体验进行反馈,造成用户体验性较差。

(2) 答疑的方式比较单一。以上介绍的几种答疑模块在设计答疑方式时比较单一,不够丰富,要么是“学生提问-教师回答”的模式,要么就是“关键词匹配数据库”的模式进行答疑,答疑方式比较单一,无法给学生提供多样性的选择。然而,在远程教育模式下,一个优秀的智能答疑系统应该不仅仅能够让系统通过

自动检索来匹配问题和答案,并且能够在学生有需要的情况下,让教师对学生的问题作出进一步的解答和说明。

纵观国内和国外的智能答疑模块,绝大部分的在线答疑都不可置否地存在些许不是很到位的问题和不足,鉴于此,研究并设计了远程教育模式下智能答疑系统,来完善现有大部分答疑系统所存在的不足。

1.2 研究目的和意义

1.2.1 研究目的

通过阅读大量国内外相关文献,研究现有市场上所有答疑系统的大致分类,分析这些答疑系统所存在的一些共性的缺点和不足。有效结合了“学生提问-教师回答”模式和“关键词匹配数据库”模式,综合运用了数据挖掘、中文分词、动态检索、自然语言处理等计算机领域的相关技术^[6-7],实现了FAQ库自动更新、自动查找匹配问题答案、关键词提取、自然语言解析等功能,力求研究、设计并开发出一套可靠、安全、易用、简便的远程教育智能答疑系统,使得答疑系统不仅仅能够通过自动检索来匹配问题和答案,并且能够在学生有需要的情况下,让教师对学生的问题作出进一步的解答和说明。

1.2.2 研究意义

为了更好地推广践行“互联网+”理念,远程教育模式作为网络教育及应用分支中一个新兴的热点问题,打破传统束缚,摆脱教育领域在时间、空间上的局限,为教师和学生提供了双重便利。研究、设计并实现远程教育系统,其智能答疑模块的智能化程度早已成为不可不考虑的重要因子。而且,伴随着数据挖掘、中文分词、动态检索、自然语言处理等计算机相关技术的不断发展,远程教育模式终将成为教育改革的必然趋势,所以智能答疑系统具有非常广阔的应用前景。

1.3 远程教育系统研究概述

1.3.1 发展历程和现状

学术界对远程教育进行研究的学者很多,而且近几年出现井喷之势,不同学者对远程教育的理解都不太一样,对远程教育所下的定义也不太一样。加利福尼亚州远程教育计划,简称CDLP,它对远程教育的理解是这样的:“远程教育,简称DL,其主要负责连接教育资源与学生之间的联系。”

远程教育的发展历程与计算机技术的发展历程比较同步,它起源于欧洲部分

发达国家以及美国。当时这些发达国家中都存在大量无法到当地学校就读的人,获得教育学习的机会不多,这些国家就开放当时世界上最先进的邮政系统让他们可以异地学习,这样使得家住偏远地方、不能去学校上学以及身体残疾的人可以远程获得教育。但当时这样模式应用不是很广泛,只在少数几个发达国家中应用,而且教育成本相对较高^[8-9]。

到了 19 世纪 80 年代,伊利诺伊卫斯理大学开发了第一套真正意义上的远程教育系统,并将这套系统用于对本校本科毕业生、硕士毕业生进行学位授予,使得他们不需要实地出席学位授予大会即可获得各自的学士和硕士毕业学位。至此,这种远程教育系统引起美国甚至全世界的关注,直到 19 世纪 90 年代,肖托夸运动又快速推动了这种远程教育的在线网络教育形式。

进入 20 世纪后,远程教育的发展遇到了新的机遇,20 世纪 20 年代出现的广播技术以及 20 世纪 40 年代出现的电视技术都为远程教育更新了崭新的通信形式。从事教育领域的人员可以利用广播和电视技术远程播放教育音频节目和视频节目,打破传播的学习模式,为无数潜在学生人员提供了良好的学习机会。

此外,在 20 世纪初期,长途电话系统发展迅猛,大大增加了参加远程教育学习的学习群体,但是相比于广播和电视技术,长途电话系统相对来说没有发挥非常重要的作用。后来,一些发达国家开发或引进了各自的电话会议技术,利用电话会议技术,教师和学生之间可以实时进行语音交谈,不管他们处于何地,都能很好地保证通话信息没有传输延迟。

除了邮政系统、广播技术、电视技术、长途电话技术、电话会议技术外,随后远程教育在发展过程中仍更新了不同的通信技术,来逐步提高教师和学生之间进行交流沟通的能力,逐步降低交流沟通所需成本。尤其是 20 世纪末计算机的出现,人们开始通过电话线连接计算机,网络的世界让师生之间的沟通交流畅通无阻,教师和学生只需要坐在会议室内就可以通过计算机进行通信和学习。计算机技术、网络通信技术的出现,教师和学生可以实时动态交换视频、音频、图片、问题等电子数据,使得在远程教育通信技术应用上,邮政系统、广播技术、电视技术、长途电话技术、电话会议技术等被逐步减少使用,甚至被淘汰。

1.3.2 系统的特点

(1) 可靠且安全。对于远程教育系统来说,可靠性和安全性是极其重要的,

Degree papers are in the “[Xiamen University Electronic Theses and Dissertations Database](#)”.

Fulltexts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.